The "Dual" and "Tridual" Dynamic Factor Analysis

Renato Coppi and Pierpaolo D' Urso
University of Rome "la Sapienza", Department of Statistics
P.le A. Moro, 5-00185, Rome, Italy
renato.coppi@uniroma1.it; pierpaolo.durso@uniroma1.it

1. "Dual" and "Tridual" Dynamic Factor Analysis

Dynamic Factor Analysis (DFA), originally introduced by Coppi and Zannella (1978) and further developed by Corazziari (1999) and Coppi, D' Urso (2000), is an exploratory methodology for analyzing 3-way arrays of the type "Units×Variables×Times", where the variables are quantitative. Let us denote by x_{ijt} the generic element of this array related to unit i (i=1,I), variable j (j=1,J), time t (t=1,T). In its 3-dimensional form this array is represented by $\mathbf{X}(I,J,T)$. However, two of the indices may be collapsed into one dimension in 3 different ways. The original proposal (Direct approach) is based on pooling together units and times along the rows, while associating the variables to the columns of a 2-way array ($\mathbf{X}(IT,J)$). Two new DFA approaches can be defined. The first one, referring to structure $\mathbf{X}(JT,I)$, is called "Dual", emphasizing the exchange of roles between units and variables, w.r.t. the Direct approach. The second one, concerning structure $\mathbf{X}(IJ,T)$, is denominated "Tridual", pointing out its twofold duality w.r.t. each of the previous approaches. Before applying any DFA model, the row data are normalized as follows: $z_{ijt} = z_{ijt} / \bar{x}_{.j.}$,$\forall$ (i,j,t), where $\bar{x}_{.j.}$ denotes the overall mean of variable j. In what follows we refer to the normalized arrays $\mathbf{Z}(I,J,T)$, $\mathbf{Z}(IT,J)$ etc. The general methodological bases of DFA, common to all approaches and models, are: 1) the decomposition of the overall variation of $\mathbf{Z}(I,J,T)$ into 3 components; 2) the modelisation and analysis of these components by means of a joint use of Singular Value Decomposition (SVD) and Regression analysis w.r.t. time (when applicable). In the Direct, Dual and Tridual approaches the overall variation is respectively measured by the covariance matrices $\mathbf{S}$, $\mathbf{P}$, $\mathbf{Q}$. In particular $\mathbf{P} = \left\{ \frac{1}{JT} \sum_{j,t} (z_{ijt} - \bar{z}_{i..})(z_{i'jt} - \bar{z}_{i'..}) \right\}_{i,i'=1,I}$ (Dual case), $\mathbf{Q} = \left\{ \frac{1}{IJ} \sum_{i,j} (z_{ijt} - \bar{z}_{..t})(z_{ijt'} - \bar{z}_{..t'}) \right\}_{t,t'=1,T}$ (Tridual case). $\mathbf{P}$ and $\mathbf{Q}$ are the Proximity Operators between units and between times, respectively.

<u>Dual case</u>. The basic decomposition of $\mathbf{P}$ is: $\mathbf{P} = \mathbf{P}^*_J + \mathbf{P}^*_T + \mathbf{P}_{JT}$, where $\mathbf{P}^*_J = \left\{ \frac{1}{J} \sum_{j} (\bar{z}_{ij.} - \bar{z}_{i..})(z_{i'j.} - \bar{z}_{i'..}) \right\}_{i,i'=1,I}$ measures the "synthetic structure" of the variable, independently of time; $\mathbf{P}^*_T = \left\{ \frac{1}{T} \sum_{t} (\bar{z}_{i.t} - \bar{z}_{i..})(\bar{z}_{i't} - \bar{z}_{i'..}) \right\}_{i,i'=1,I}$

represents the variation due to the "average time evolution" of the units (over the whole set of variables), and $\mathbf{P}_{JT} = \left\{ \frac{1}{JT} \sum_{j,t} (z_{ijt} - \bar{z}_{ij.} - \bar{z}_{i.t} - \bar{z}_{i..})(z_{i'jt} - \bar{z}_{i'j.} - \bar{z}_{i'.t} - \bar{z}_{i'..}) \right\}_{i,i'=1,I}$ is a measure of the "residual" variation due to the "differential evolution" of the variables (interaction between variables and times). The pooled operator $\overline{\mathbf{P}}_T = \mathbf{P}_J^* + \mathbf{P}_{JT}$ measures the global "structure" of the variables (including their differential evolution), while $\overline{\mathbf{P}}_J = \mathbf{P}_T^* + \mathbf{P}_{JT}$ measures the global dynamics of the system.

Model 1 (in the dual approach) consists of the SVD of $\overline{\mathbf{P}}_T$ (providing components scores and factorial trajectories of each variable) and the polynomial regression analysis of $\bar{z}_{i.t}$ w.r.t. time (accounting for the variation measured by $\mathbf{P}_T^*$).

Model 2 is based on the SVD of $\mathbf{P}_J^*$ and time regression analysis applied to $\overline{\mathbf{P}}_J$. The information provided by the two Models summarizes the structure and the dynamics of the system, from the viewpoint of proximities between units. This can be matched with the Direct approach which conveys the analogous information, in the perspective of proximities between variables.

Tridual case. Here we look at the proximities (similarities) between times. In obvious notation, the basic decomposition is $\mathbf{Q} = \mathbf{Q}_I^* + \mathbf{Q}_J^* + \mathbf{Q}_{IJ} = \overline{\mathbf{Q}}_I + \mathbf{Q}_I^* = \overline{\mathbf{Q}}_J + \mathbf{Q}_J^*$. All of these operators are analyzed by SVD providing additional information on the role of time w.r.t. the units and the variables.

A strategy of analysis for the data in $\mathbf{X}(I,J,T)$ can be devised on the above grounds, integrating the Direct DFA approach with the Dual and Tridual approaches. An application to the study of the economic evolution of the Italian industrial system in the last 15 years is illustrated.

REFERENCES

Coppi, R., D'Urso, P. (2000) The Dual Dynamic Factor Analysis Model, *24th Conference of GfKl*, 15-17 March, Passau.

Coppi R., Zannella F. (1978) L'Analisi Fattoriale di una Serie Temporale Multipla relativa allo Stesso Insieme di Unità Statistiche, *XXIX Meeting of the Italian Stat. Soc.*, Bologna, 61-79.

Corazziari, I. (1999), Dynamic Factor Analysis, in Vichi, M., Opitz O. (Eds.), *Classification and Data Analysis. Theory and Application*, Springer-Verlag, Heidelberg, 171-178.

RESUME

Les bases méthodologiques de l'Analyse Factorielle Dynamique sont illustrées:1) La décomposition de la matrice des covariances de un array concernant l'observation de J variables quantitatives sur I individus en T temps successifs; 2) La modélisation de ces composantes moyennant l'application conjointe de la Décomposition en valeurs singulières et de la Régression par rapport au temps. Les approches "dual" et "tridual" sont définies dans ce contexte et leur intégration avec l'approche directe est considérée.